AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 108:147–159 (1999)

Population Substructure and Isolation by Distance in Three Continental Regions

ELISE ELLER*
Department of Anthropology, The Pennsylvania State University, University Park, Pennsylvania 16801
Department of Anthropology, University of Utah, Salt Lake City, Utah 84112

KEY WORDS short tandem repeats (STRs); principal components analysis (PCA); Mantel test

ABSTRACT Isolation by distance and divergence from a shared population history are two sources of population substructure. Isolation by distance erases population history as populations approach migration-drift equilibrium, while diverging populations descended from a single ancestral population will accumulate genetic differences with time. Here I investigate how much of the worldwide genetic diversity from Jorde et al.'s ([1997] Proc. Natl. Acad. Sci. U S A *94:*3100–3103) 60 tetranucleotide short tandem repeat (STR) data can be explained by isolation by distance. I use Slatkin's measure of population substructure, R_{ST}, principal components analyses, and Mantel tests to investigate the pattern of genetic diversity at both the intercontinental and intracontinental levels. Geographic distance accounts for almost 60% of worldwide interpopulation genetic relationships. Within continents, the correlations are less, although not significantly so because of wide confidence intervals. These results suggest that populations have not reached migration-drift equilibrium and that there is information in STR data to reconstruct population history.

The level of population substructure worldwide is consistent with previous observations, but at the intracontinental level substructure is less. When one examines diversity against distance from the centroid, one sees excess heterozygosity in Africa, a pattern also noted by Stoneking et al. ([1998] Genome Research *7:*1061–1071). A larger effective population size in Africa could explain the excess diversity. Greater gene flow in Africa is an unlikely explanation because the African R_{ST} value is slightly larger than the Asian and European R_{ST}s, pointing to less gene flow and greater substructure among African populations. Furthermore, there are differences in patterns between heterozygosity and allele size variance. Heterozygosity has a higher correlation with distance from the centroid than does allele size variance, and this may reflect demographic history. Kimmel et al. ([1998] Genetics *148:*1921–1930) have shown that after a population expansion heterozygosity returns to equilibrium more quickly than does allele size variance. The contrasting patterns between heterozygosity and allele size variance may reflect different times after an expansion. However, simulations and further work need to be done to more thoroughly investigate the possibility that these data reflect population expansions. Am J Phys Anthropol 108:147–159, 1999.

Grant sponsor: National Science Foundation.

*Correspondence to: Elise Eller, Department of Anthropology, 102 Stewart Building, University of Utah, Salt Lake City, UT 84112. E-mail: elise.eller@anthro.utah.edu

Received 11 December 1997; accepted 17 October 1998.

From a variety of genetic and craniometric data, researchers have noted that most variation is within populations and not among populations. Some population substructure exists, on the order of 10–15% (Barbujani et al., 1997; Bowcock et al., 1991; Cavalli-Sforza et al., 1988; Deka et al., 1995; Jorde et al., 1995; Lewontin, 1972; Nei and Livshits, 1990; Relethford and Harpending, 1994), but most population geneticists would agree that there is no biological basis for race, sensu Mayr (1963), in which race defines a discontinuous set of individuals within a species (see also Barbujani et al., 1997). However, in a principal components map based on 60 tetranucleotide loci published in Jorde et al. (1997), the 15 populations surveyed clustered in groups reminiscent of traditional concepts of races. This clustering of populations could be an artifact of the spatial distribution of the populations chosen for study. Because the populations surveyed cluster geographically, it is not surprising that they cluster genetically as a result of isolation by distance. On the other hand, the clustering could reflect other or additional processes, such as divergence from a shared population history. In this study I compare measures of population substructure both at the worldwide level and at the continental level, present principal components analyses (PCA), and perform Mantel tests to investigate patterns of population substructure and test whether the observed patterns of genetic diversity can be explained by an isolation by distance model.

To measure population substructure, I use Slatkin's (1995) statistic R_{ST}, which is a more appropriate measure of substructure for short tandem repeats (STRs) because it incorporates allele size variance while statistics such as F_{ST} and G_{ST} do not. R_{ST} therefore makes better use of information inherent in STRs and does not assume a low mutation rate (Slatkin, 1995). The principal components analyses provides a graphical approach for displaying genetic distances between populations. Furthermore, it can be used to investigate the relationship between diversity and distance from the centroid. Harpending and Jenkins (1973) and Harpending and Ward (1982) have developed methods to explore deviations from the theoretical relationship between genetic diversity and distance from the centroid. However, PCA is a graphical and not statistical approach. Mantel tests, in contrast, are statistical tests for testing correlations between distance matrices. Using the Mantel test, I calculate the correlation between genetic distance and geographic distance while controlling for linguistic affiliation and population size to see whether the patterns of genetic diversity fit an isolation by distance model.

MATERIALS AND METHODS

Genetic data

The data are allele size frequencies of 60 tetranucleotide loci reported in Jorde et al. (1997). Samples include six sub-Saharan African populations (Biaka Pygmies, n = 5; Bushmen, n = 15; Mbuti Pygmies, n = 5; Nguni, n = 12; Sotho/Tswana, n = 21; Tsonga, n = 14), five East or Southeast Asian populations (Cambodian, n = 12; Han Chinese, n = 17; Japanese, n = 19; Malay, n = 6; Vietnamese, n = 9), and four European populations (Finns, n = 20; French, n = 20; "Northern Europeans," a sample of Caucasians collected in the Salt Lake City area and predominantly of British or Scandinavian descent, n = 70; Poles, n = 10), for a total of 255 diploid individuals or 510 haploid genomes. The samples are described more thoroughly in Jorde et al. (1995, 1997).

Analytic methods

R_{ST}. R_{ST} is Slatkin's (1995) STR-specific measure of population substructure analogous to F_{ST} where

$$R_{ST} = \frac{\frac{1}{l}\sum_{i=1}^{l}(\overline{S}_i - S_{W_i})}{\frac{1}{l}\sum_{i=1}^{l}\overline{S}_i}.$$

$\overline{S}_i$ is the mean squared pairwise differences in allele size at locus i (equivalently, twice the allele size variance at that locus) in the worldwide sample, S_{Wi} is the mean squared pairwise differences of each population averaged over all 15 populations weighted by sample size at locus i, as suggested in Sokal and Rohlf (1995), and l is the number of loci.

R_{ST} also can be calculated at the continental level. In this case, $\overline{S}_i$ is the mean squared pairwise differences within the continent, and S_{Wi} is the mean squared pairwise differences of each population averaged over all populations within that continent and weighted by sample size.

Principal components analysis. Principal components analyses were performed by standard methods (Harpending and Jenkins, 1973). I created a 60 × 15 matrix **Z** in which each element $z_{i,j}$ was the normalized mean allele size at locus i in population j:

$$z_{i,j} = \frac{\overline{p}_{i,j} - \overline{p}_i}{s_i}$$

where $\overline{p}_{i,j}$ is the mean allele size at locus i for population j, $\overline{p}_i$ is the mean allele size at locus i in the worldwide sample, and the denominator is the standard deviation of allele sizes at locus i in the worldwide sample. Principal coordinates were derived from the first two eigenvalues and eigenvectors of $\mathbf{R} = \mathbf{Z}^{\mathbf{T}}\mathbf{Z}$. The PC map can be compared to the geographic locations of the sampled populations using a matrix rotation-contraction-translation method described in Schönemann and Carroll (1970). Distances from the centroid were taken from the diagonal of **R** and corrected for bias by subtracting by $1/2n$, where n is the sample size for that population. These were plotted against heterozygosity and against allele size variance, which were calculated directly from the data and corrected for bias.

Mantel tests. Mantel tests (Smouse and Long, 1992; Mantel, 1967) were performed to test the fit of a matrix of genetic distances **D** to a matrix of geographic distances **G** while holding linguistic affiliation **L** and population size **N** constant (**D**-**G**.**L**, 1/**N**). Genetic distances between pairs of populations are Shriver et al.'s (1995) D_{SW} averaged over the 60 loci. Geographic distances were measured on a globe and reflect a best guess of the shortest path from one location to another while avoiding major geographic barriers such as oceans and the Himalayas. Although Morton (1973) noted that empirically genetic distance increases logarithmically with geographic distance, in these analyses there was no significant difference between using geographic distances vs. the natural logarithm of geographic distances at such a large geographic scale (not shown), and I chose to use geographic distances and not their natural logarithms.

TABLE 1. Population substructure for 60 STR loci in 15 populations

Population	R_{ST} ± SE[a]
World	0.0918 ± 0.0145
Africa	0.0283 ± 0.0319
Asia	0.0186 ± 0.0272
Europe	0.0171 ± 0.0274
Eurasia	0.0502 ± 0.0172

[a] R_{ST} is Slatkin's (1995) measure of population substructure for STRs (see Materials and Methods). Standard errors were generated with 5,000 bootstraps.

The Mantel tests also incorporated language affiliation (**L**) and population size (**N**). The matrix of language affiliation is a binary matrix where language affiliation between a pair of populations is 1 if there was no affiliation at the phylum level and 0 if there is a linguistic affiliation. Ruhlen's (1992) book was used to determine language affiliation. Population size was also considered since the effects of genetic drift are greater in smaller populations. Each element of the population size matrix $n_{i,j}$ is the harmonic mean $[1/n_i + 1/n_j]^{-1}$ of the pair of populations i and j to emphasize the greater effect of smaller population sizes. For ethnic groups that more or less coincide with nation-states (e.g., Vietnamese, Poles), I used current census data from a government web site that breaks down countries by ethnic group (⟨http://www.odci.gov/cia/publications/factbook/country-frame.html⟩). For ethnic groups that do not coincide with national boundaries (e.g., Nguni, Bushmen), population size estimates were found in Murdock's (1959) book. All population sizes were crude estimates in that they are current census sizes and do not include population size changes over time. However, the estimates give a rough basis for comparison.

RESULTS

R_{ST} values

R_{ST} values are shown in Table 1. The amount of substructure found worldwide, 0.0918, is consistent with other published

TABLE 2. Genetic diversity for 60 STR loci in 15 populations

Population	2n[a]	Variance[b]
World	510	4.7934
Africa	144	4.6254
Biaka Pygmy	10	5.1245
Bushman	30	3.9375
Mbuti Pygmy	10	4.8210
Nguni	24	4.8397
Sotho/Tswana	42	4.4850
Tsonga	28	4.4938
African mean (six populations)[c]		4.4944
Asia	126	4.5248
Cambodian	24	4.7343
Chinese	34	4.5098
Japanese	38	4.1208
Malay	12	4.2291
Vietnamese	18	4.7403
Asian mean (five populations)[c]		4.4404
Europe	240	4.2992
Finns	40	3.8982
French	40	4.2994
Northern Europeans	140	4.3418
Poles	20	3.8970
European mean (four populations)[c]		4.2257
World mean (fifteen populations)[c]		4.3531

[a] 2*n* is the haploid sample size, or twice the number of individuals.
[b] Bias-corrected allele size variance, which is one-half the mean squared pairwise difference used in Slatkin's (1995) R_{ST} statistic.
[c] All means (world, Africa, Asia, Europe) are weighted by sample size as suggested in Sokal and Rohlf (1995, p. 182).

measures of population substructure but is larger than Jorde et al.'s (1995) G_{ST} estimate of 0.034 from the same dataset. This difference most likely reflects the different approaches of measuring population substructure: R_{ST} incorporates allele size variance, while G_{ST} uses heterozygosity. R_{ST} values within continents were much lower than the overall R_{ST} and ranged from about 1.7–2.8%. Africa has the most population substructure, while Asia and Europe have less. This pattern has been observed in other studies (see Jorde et al., 1998; Relethford, 1995; references therein). However, the 95% confidence intervals overlap, and these estimates of intracontinent substructure are not significantly different from each other or from the worldwide estimate of R_{ST}.

Furthermore, African populations tend to have more variation than non-African populations. This pattern is consistent with the findings in many other studies (for two good reviews, see Relethford, 1995; Jorde et al., 1998). Allele size variances averaged over the 60 loci ranged from 3.90 (Finns and Poles) to 5.12 (Biaka) (see Table 1). When one looks at the individual loci (not shown), however, there is no pattern; a population with a low allele size variance at one locus may have an intermediate or high allele size variance at another locus. Although Africans populations were not consistently more variable than populations from other continents, perhaps because of the populations sampled and the small sample sizes involved, usually at least two of the six African populations were in the top three largest allele size variances. Another analysis described in Jorde et al. (1997) showed statistically that variability is greater in Africa than in the other two continents.

Principal components analyses

Figure 1a shows the principal coordinates plot. In Figure 1b, the principal coordinates plot is rotated, expanded, and translated with Schönemann and Carroll's (1970) least-squares algorithm to fit a map plotting the geographic locations sampled. Figure 1b compares graphically the genetic relationships from the PC map and the geographic relationships among the sampled populations. Although genetics and geography broadly mirror each other, Figure 1b reveals incongruities that arguably reflect demographic history, but note that this method is purely descriptive and does not specify which deviations of the genetic map from the geographic map are significant. One can create all sorts of post-hoc explanations for the deviations of the principal coordinates from their geographic locations, including small effective population size and thus large effects of genetic drift among the Pygmy groups; migration of Bantu-speaking groups from the north and west; and higher amounts of gene flow between European and Asian populations or recent migrations from the Eurasian steppes into Europe and Asia. However, since these are post-hoc speculations and the PCA method does not indicate which deviations are significant, such stories should be taken provisionally.

At the continental level, there are differences between the African PC plot and the Asian and European PC plots (Fig. 2). The African plot shows a clustering of the Bantu groups and the Bushmen with the two Pygmy groups distant from this cluster and from

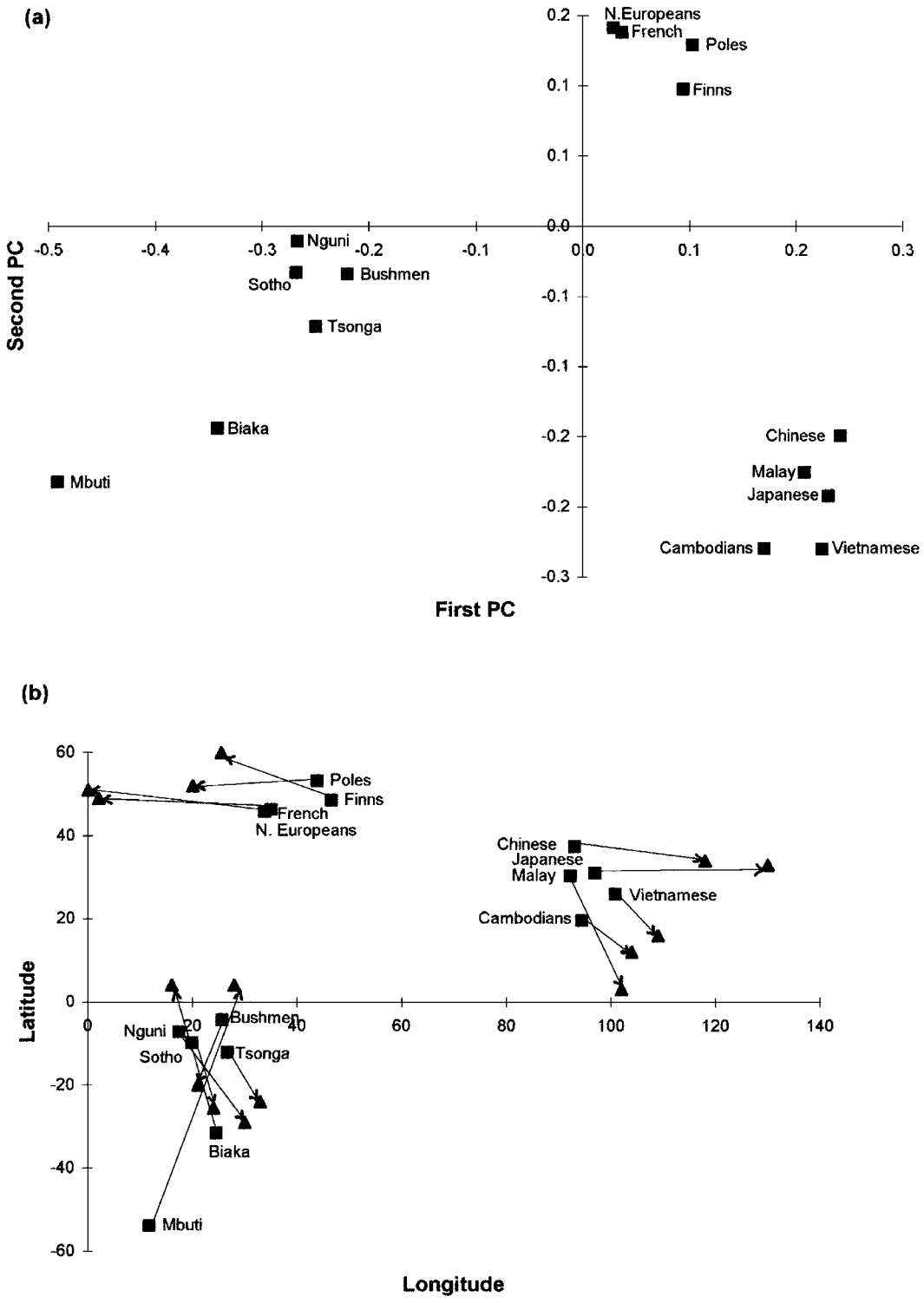

Fig. 1. **a:** Plot of the first two principal coordinates of the Jorde et al. (1997) data set. **b:** The PC plot rotated, expanded, and translated to fit the geographic locations of the sampled populations according to the Schönemann and Carroll (1970) least-squares algorithm. The rotated principal components locations (squares) are mapped (→) onto their corresponding geographic locations (triangles). See text for details.

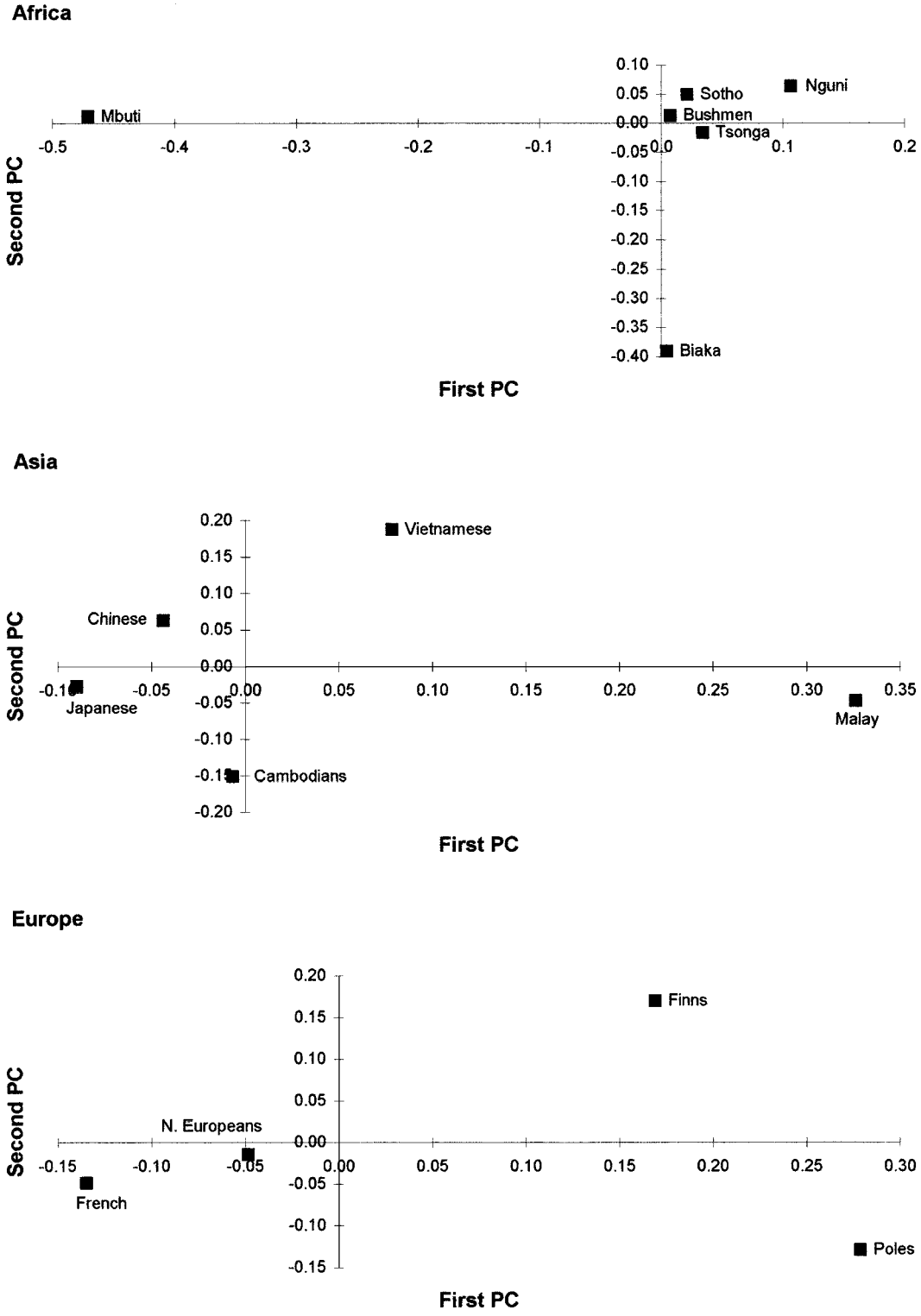

Fig. 2. Principal coordinates plot for Africa , Asia , and Europe

each other. In the Asian and the European PC plots, the populations are more dispersed and show no clustering. However, when the Asian and European populations are analyzed together in a Eurasian PC plot (not shown), clustering reappears. The first component separates the European populations from the Asian populations, while the second component distinguishes the Malay, Finns, and Poles from the others.

Plots of the genetic diversity (heterozygosity or allele size variance) against distance from the world centroid are shown in Figure 3. Theory predicts that the more distant a population is from the centroid, the smaller its heterozygosity and the less its variance in the absence of confounding effects such as gene flow from an external source and assuming equal effective population size (Harpending and Ward, 1982). Harpending and Ward (1982) derived mathematically the relationship of heterozygosity vs. distance from the centroid, but they noted that this relationship applies to other measures of genetic variation. Interestingly, in the heterozygosity vs. distance from the centroid plot, the Asian and European populations lie along the theoretical regression line, while the African populations show an excess of diversity. This pattern was also found in Stoneking et al.'s (1998) analysis of eight *Alu* loci. This pattern of excess diversity is not as evident in the graph of allele size variance vs. distance from the centroid (Fig. 3b). Some African as well as some Asian populations show excess diversity, while all European populations, two Asian populations, and the Bushmen have less than the expected allele size variance. In addition, the two Pygmy groups are extremely distant from the centroid and show an excess of both heterozygosity and allele size variance.

An interesting pattern emerges when the centroid is not the world centroid but the African centroid. In this model, Africa is considered to be the genetic center, while Asia and Europe are peripheral to Africa. In Figure 4a, heterozygosity and distance from the centroid are closely correlated, although the Asian and European populations as well as the Biaka have higher heterozygosities than those predicted from the theoretical regression line. Similarly, in the allele size variance graph, most populations show an excess of diversity, but the correlation between allele size variance and distance from the centroid is not as strong as that between heterozygosity and distance from the centroid. Therefore, we see two interesting results from this graph: a consistent excess of Asian and European heterozygosity and allele size variance compared to the theoretical regression line and a weaker correlation of allele size variance to distance from the centroid compared to that of heterozygosity and distance from the centroid. The weaker correlation of allele size variance may be a signal of a population expansion (Kimmel et al., 1998). This will be discussed further below.

In the genetic diversity vs. distance from the centroid plots at the continental level (Fig. 5), no real pattern is apparent. A population with high heterozygosity does not necessarily have a high allele size variance. The lack of correlation between genetic diversity and distance from the continental centroid is not the pattern one sees at the worldwide level with an African centroid, where there is a correlation between diversity and distance from the centroid, especially for heterozygosity.

Mantel tests

Table 3 shows the partial correlation coefficients for genetic distance vs. geographic distance while controlling for linguistic affiliation and population size (**D**-**G**.**L**, 1/**N**). Overall, r is 0.767. This value is not very different from the correlation coefficient within Africa (r = 0.623) but quite different from the correlation coefficients for Asia (r = -0.037) and Europe (r = -0.910). Only the worldwide and African correlation coefficients show significantly positive values. The Asian and European correlation coefficients are negative (significantly so in the case of Europe); however, the 95% confidence intervals are extremely wide, and the sample sizes are five and four, respectively. When the Asian and European populations are lumped into a Eurasian cluster, the correlation coefficient is once again significantly positive (r = 0.877). This correlation might be an effect of having large, intercontinent distances in the

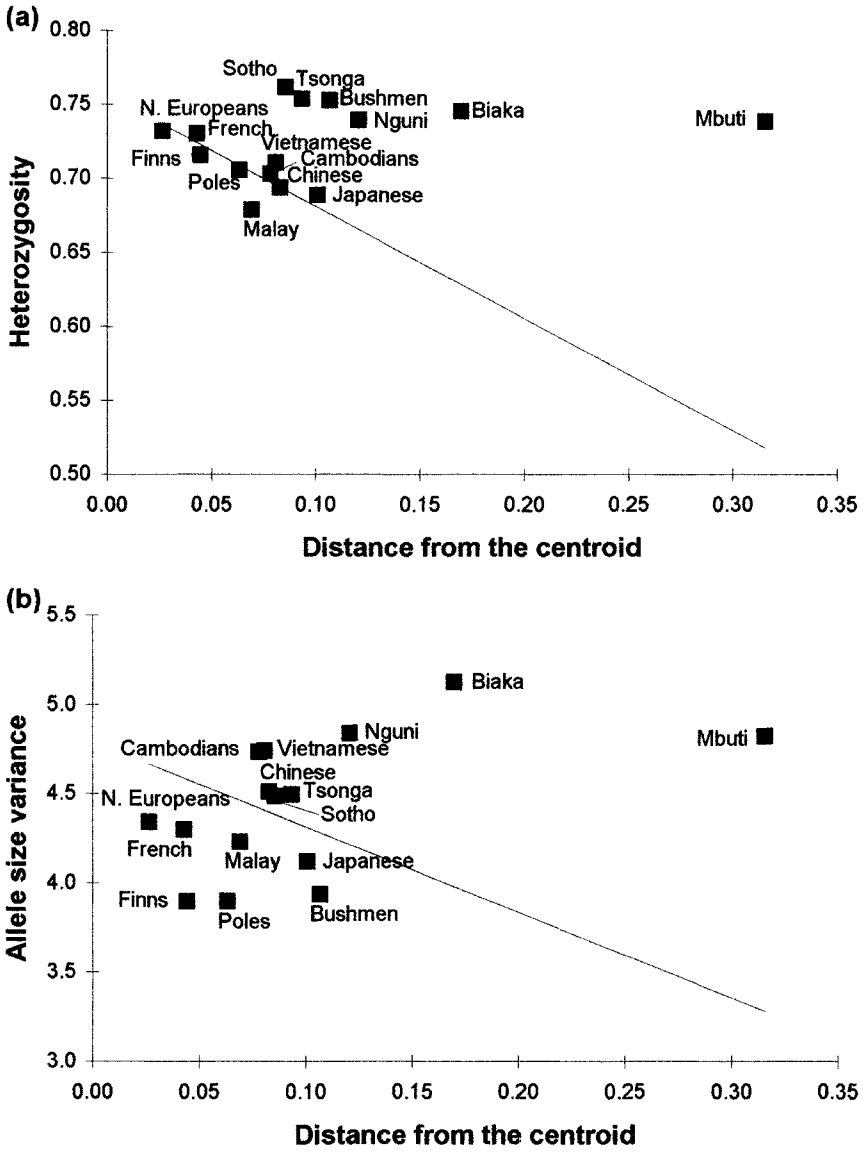

Fig. 3. Heterozygosity vs. distance from the centroid (**a**) and allele size variance vs. distance from the centroid (**b**) for the worldwide sample. Solid lines are the theoretical regression lines as described in Harpending and Ward (1982): $E[h_i] = H(1 - r_i)$ where h_i is heterozygosity for population i, H is the overall observed heterozygosity, and r_i is the distance from the centroid for population i; $E[var_i] = s^2(1 - r_i)$ where var_i is the allele size variance for population I, s^2 is the overall observed allele size variance, and r_i is distance from the centroid for population i.

geographic distance matrix as well as smaller, intracontinent distance values, thus artificially inflating the correlation. However, if one does not reject this correlation coefficient as a spurious result, then this result suggests that one should not compare each continental region to others but Africans vs. non-Africans. This idea should be tested with a greater number of sampled populations.

DISCUSSION

What do these results say about genetic substructure in humans today? One possible

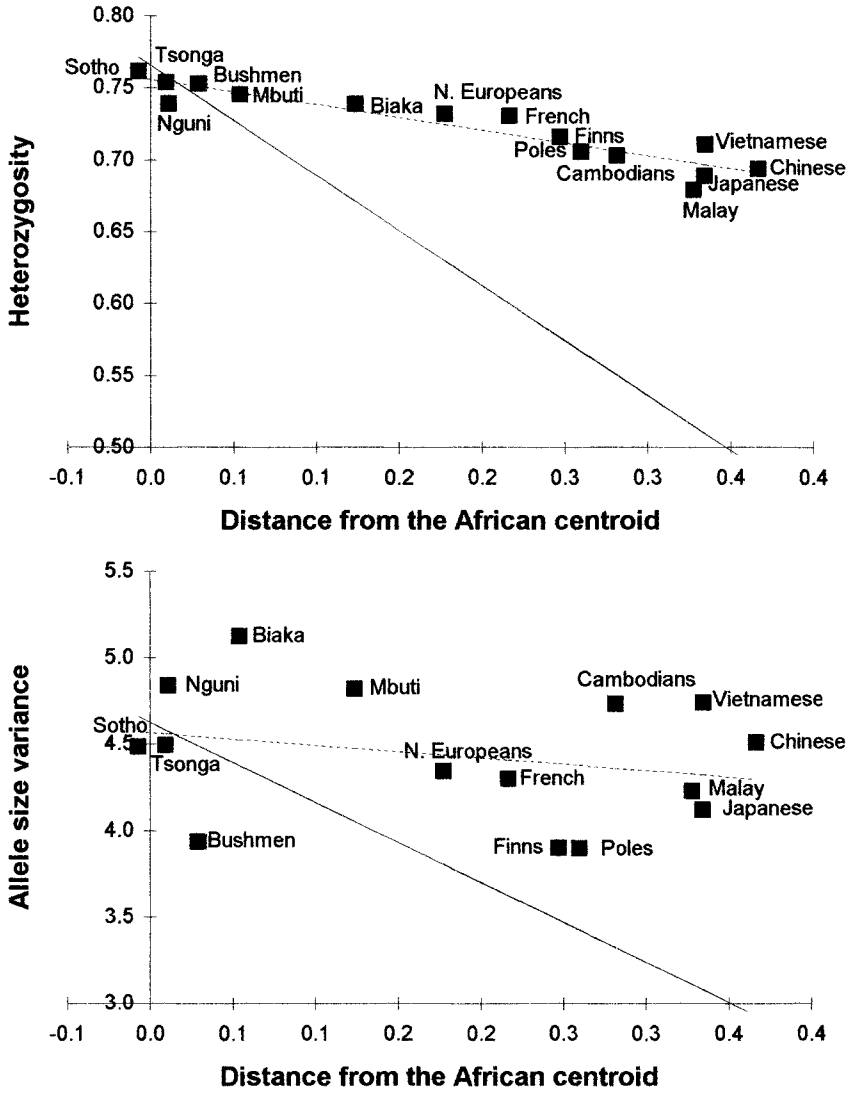

Fig. 4. Heterozygosity vs. distance from the African centroid and allele size variance vs. distance from the African centroid for the worldwide sample. Solid lines are the theoretical regression lines as described above, except H and s^2 are the observed heterozygosity and allele size variance in Africa. Dashed lines are the empirical regression lines. For heterozygosity, $h_i = 0.7559 - 0.1772r_i$, Pearson's $r = -0.936$, and for allele size variance $var_i = 4.5668 - 0.7343r_i$, Pearson's $r = -0.267$.

cause of population substructure is population bifurcation, in which populations diverge with little or no subsequent gene flow between the branches. This model would predict more substructure among continental regions than is observed today, and this sort of population history has been rejected for humans (Weiss and Maruyama, 1976; Templeton, 1997), although it is often implicit in many research articles. However, levels of population substructure will be minimal if levels of gene flow among continental regions are relatively high. Another possible cause is an isolation by distance model in which gene exchange occurs more frequently between geographically closer populations, and this gene flow erases past population history once equilibrium has been

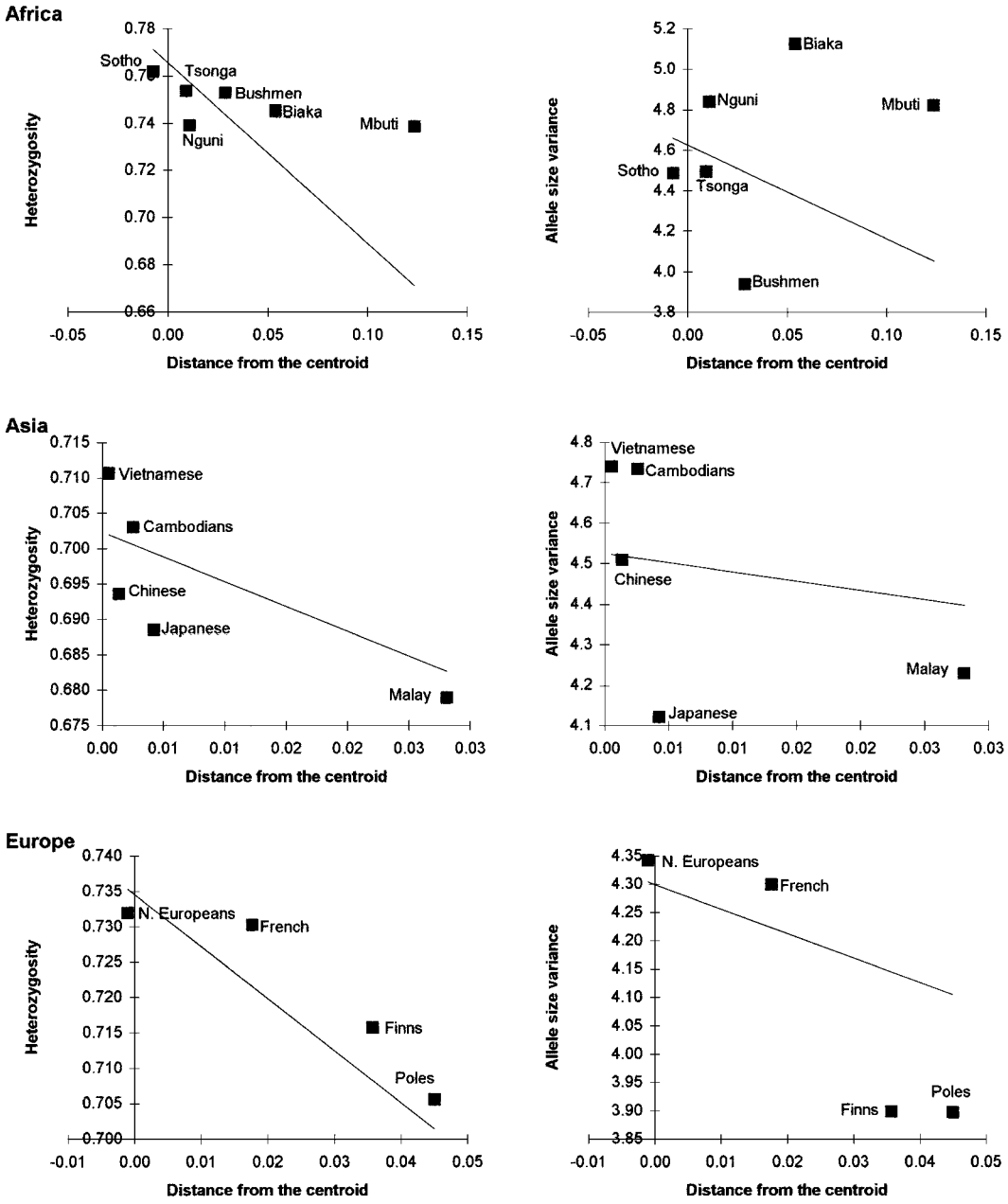

Fig. 5. Heterozygosity vs. distance from the centroid (**left**) and allele size variance vs. distance from the centroid (**right**) for Africa (**top**), Asia (**middle**), and Europe (**bottom**). Solid lines are the theoretical regression lines as described above, except H and s^2 are the observed heterozygosities and allele size variances for each continent.

reached. In this case, genetic distance correlates with geographic distance.

The evidence from these 60 tetranucleotide loci suggests that the isolation by distance model better explains contemporary population substructure in humans. Correlations between genetic distance and geographic distance, while controlling for lin-

*TABLE 3. Partial correlation coefficients of intrapopulation genetic distances and geographic distances while holding language affiliation and population size constant (**D**−**G**.**L** 1/**N**)*[a]

Population	r[b]	p[c]	95% CI[d]
World	0.767	0.001	0.693–0.823
Africa	0.623	0.009	0.353–0.890
Asia	−0.037	0.520	−0.596–0.704
Europe	−0.901	0.010	−1.000–1.000
Eurasia	0.877	0.001	0.799–0.933

[a] **D**, matrix of genetic distances; **G**, matrix of geographic distances; **L**, binary matrix of linguistic affiliation; **N**, matrix of population sizes. See text for details.
[b] Partial correlation coefficient as described in Smouse and Long (1992).
[c] Probability that the estimated correlation coefficient does not lie within the null distribution of no structure (Smouse and Long, 1992). One thousand permutations performed.
[d] Generated from 1,000 bootstraps.

guistic affiliation and population size, are high at the worldwide level and within Africa and within Eurasia. Geographic distance, linguistic affiliation, and population size explain nearly 60% of the genetic diversity worldwide. Similarly high values (nearly 40% and 77%, respectively) are true for Africa and Eurasia, although the high Eurasian value could be a statistical artifact, as discussed above (Results). Correlation coefficients within Asia and within Europe, on the other hand, are not significantly positive. This could be a statistical artifact of having little diversity within Asia and Europe and having so few populations sampled in these two continents, or it might reflect demographic processes. Either there have been recent migrations into Europe and Asia (e.g., from the Eurasian steppes) and not enough time has occurred for populations to differentiate, or gene flow was sufficiently high within Asia and Europe (but not Africa) to erase any preexisting substructure and prevented any positive correlation between genetic distance and geographic distance.

The fact that R_{ST} at the worldwide level is greater than at the continental level, although these estimates are not significantly different by bootstrapping, also has demographic implications. There is less substructure within continents than among continents, and this suggests that there has been less time for substructure to develop within continents than among continents or there has been much more gene flow within continents than among continents—enough gene flow to erase any preexisting substructure.

Finally, the diversity vs. distance from the centroid graphs reveals interesting patterns that have implications about human demographic history. First, the graph of heterozygosity vs. distance from the centroid (Fig. 3a) reveals an excess of heterozygosity among the African populations, while the Asian and European populations lie close to the theoretical regression line. This pattern of excess heterozygosity among African populations was also reported in Stoneking et al. (1998). They suggest that this excess heterozygosity is a result of larger effective population size in Africa. They noted that another possible explanation was more gene flow among African populations but rejected this alternative because G_{ST} was greater in Africa than in other regions of the world. This argument can be applied to these STR data as well: R_{ST} is slightly larger in Africa than in Asia or Europe, and therefore there cannot be more gene flow among African groups but less, thus increasing population substructure. Excess heterozygosity, therefore, must be a result of larger effective population size in Africa. The graph of allele size variance vs. distance from the centroid (Fig. 3b) suggests a similar pattern of excess diversity, but it is not as consistent.

Second, when the African centroid is used (Fig. 4a), the observed heterozygosities lie in a straight line, although not along the theoretical regression line based on mean African heterozygosity. In other words, while the Asian and European populations and the Biaka have excess diversity compared to expectation, there is still a strong linear relationship between heterozygosity and distance from the African centroid, as predicted by Harpending and Ward (1982). The relationship between allele size variance and distance from the African centroid is similar although the correlation is not as strong. These patterns suggest that Africa is a core area and that Europe and Asia are peripheral. This is not a surprising result, since Africa is known to be the birthplace of the hominid lineage and there has been at least one great migration out of Africa 1.8 million years ago, if not additional, later dispersals out of Africa as proponents of the recent African origin model argue. However, Asia

and Europe are not as peripheral as the theoretical regression line would predict.

Third, the diversity vs. distance from the centroid graphs reveals an interesting contrast between the two measures of diversity used in these analyses, heterozygosity and allele size variance. This contrast is most evident in the distance from the African centroid graphs (Fig. 4) but is also noticeable in the graphs at the continental level (Fig. 5). While there is a strong linear relationship between heterozygosity and distance from the African centroid, as noted above, the correlation between allele size variance and distance from the African centroid is not as strong. These contrasting patterns could reflect demographic history. Kimmel et al. (1998) have shown that when a population expansion has occurred, estimates of population size based on homozygosity increase more quickly than estimates of population size based on allele size variance. Heterozygosity returns to equilibrium more quickly than the variance. These two contrasting patterns, therefore, could be reflecting different times in demographic history. This prospect warrants further investigation, and I have begun simulations and other analyses to address this question.

It is important to keep in mind a few caveats, however. First, sample sizes of populations such as the Biaka and Mbuti Pygmies and the Malays are small. Also, a wide geographic range was sampled with only 15 populations; this sampling only sparsely samples the genetic and linguistic variation throughout the world. More populations within each continental region need to be surveyed to capture more fully the genetic variation in each continent and to decrease bootstrapped within-continent confidence intervals of R_{ST} and correlation coefficients. In addition, because populations cluster within each continent, the distribution of geographic distances includes many shorter distances within continents and many longer distances among continents, but there are few intermediate distances. The larger distances potentially have greater influence in the Mantel test and could inflate the correlation between genetic distance and geographic distance at the worldwide level but not at the continent level. This effect could explain the high correlation for Eurasia when the results of the Mantel tests of Asia and Europe separately are negative. Therefore, we should take these results as preliminary but indicative that much of our population history—in STRs at least—has been lost by local gene flow but there is still enough information in the data to retrieve human population history.

ACKNOWLEDGMENTS

I thank Lynn Jorde for providing his data. Thanks go also to Henry Harpending for advice and helpful discussions and to three anonymous reviewers for comments and suggestions. This work was partially supported by a National Science Foundation predoctoral fellowship.

LITERATURE CITED

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. 1997. An apportionment of human DNA diversity. Proc Natl Acad Sci U S A 94:4516–4519.

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci U S A 88:839–843.

Cavalli-Sforza LL, Piazza P, Menozzi A, Mountain JL. 1988. Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data. Proc Natl Acad Sci U S A 85:6002–6006.

Deka R, Li J, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R. 1995. Population genetics of dinucleotide $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56:461–474.

Harpending HC, Jenkins T. 1973. Genetic distances among southern African populations. In: Crawford MH, Workman PL, editors. Methods and theories of anthropological genetics. Albuquerque, NM: University of New Mexico Press. p 177–199.

Harpending HC, Ward RH. 1982. Chemical systematics and human populations. In: Nitecki MH, editor. Biochemical aspects of evolutionary biology. Chicago: University of Chicago Press. p 213–256.

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak P, Carpenter KD, Soodyall H, Jenkins T, Rogers AR. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538.

Jorde LB, Rogers AR, Bamshad MJ, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC. 1997. Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci U S A 94:3100–3103.

Jorde LB, Bamshad MJ, Rogers AR. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. Bioessays 20:126–136.

Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB. 1998. Signatures of population expansion in microsatellite repeat data. Genetics 148:1921–1930.

Lewontin RC. 1972. The apportionment of human diversity. Evol Biol 6:381–398.

Mantel NA. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–218.

Mayr E. 1963. Animal species and evolution. Cambridge, MA: Belknap.

Morton NE. 1973. Isolation by distance. In: Morton NE, editor. Genetic structure of populations. Honolulu, HI: University of Hawaii Press. p 76–77.

Murdock GP. 1959. Africa: its peoples and their culture history. New York: McGraw-Hill Book Company, Inc.

Nei M, Livshits G. 1990. Evolutionary relationships of Europeans, Asians, and Africans at the molecular level. In: Takahata N, Crow JF, editors. Population biology of genes and molecules. Tokyo: Baifukan. p 251–265.

Relethford JH. 1995. Genetics and modern human origins. Evol Anthropol 4:53–63.

Relethford JH, Harpending HC. 1994. Craniometric variation, genetic theory, and modern human origins. Am J Phys Anthropol 95:249–270.

Ruhlen M. 1992. A guide to the world's languages, vol. 1: classification, 2nd ed. London: Edward Arnold.

Schönemann PH, Carroll RM. 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. Psychometrika 35:245–443.

Shriver MD, Li J, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol Biol Evol 12:914–920.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Smouse PE, Long JC. 1992. Matrix correlation analysis in anthropology and genetics. Yearbook of Physical Anthropology 35:187–213.

Sokal RR, Rohlf FJ. 1995. Biometry, 3rd ed. New York: W.H. Freeman and Company.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1998. *Alu* insertions polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Research 7:1061–1071.

Templeton AR. 1997. Out of Africa? What do genes tell us? Curr Opin Genet Dev 7:841–847.

Weiss KM, Maruyama T. 1976. Archeology, population genetics and studies of human racial ancestry. Am J Phys Anthropol 44:31–50.